ELSEVIER

Available at www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®

Journal of Multivariate Analysis 92 (2005) 227–238

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Improving on the mle of a bounded location parameter for spherical distributions

Éric Marchand[a,1] and François Perron[b,*,1]

[a] *Department of Mathematics and Statistics, University of New Brunswick, P.0. Box 4400, Fredericton, NB, Canada E3B 5A3*

[b] *Département de Mathématiques et de Statistique, Université de Montréal, P.O. Box 6128, Succursale Centre-Ville, Montréal, Que., Canada H3C 3J7*

Received 17 January 2002

Abstract

For the problem of estimating under squared error loss the location parameter of a p-variate spherically symmetric distribution where the location parameter lies in a ball of radius m, a general sufficient condition for an estimator to dominate the maximum likelihood estimator is obtained. Dominance results are then made explicit for the case of a multivariate student distribution with d degrees of freedom and, in particular, we show that the Bayes estimator with respect to a uniform prior on the boundary of the parameter space dominates the maximum likelihood estimator whenever $m \leqslant \sqrt{p}$ and $d \geqslant p$. The sufficient condition $m \leqslant \sqrt{p}$ matches the one obtained by Marchand and Perron (Ann. Statist. 29 (2001) 1078) in the normal case with identity covariance matrix. Furthermore, we derive an explicit class of estimators which, for $m < \sqrt{p}$, dominate the maximum likelihood estimator simultaneously for the normal distribution with identity covariance matrix and for all multivariate student distributions with d degrees of freedom, $d \geqslant p$. Finally, we obtain estimators which dominate the maximum likelihood estimator simultaneously for all distributions in the subclass of scale mixtures of normals for which the scaling random variable is bounded below by some positive constant with probability one.

AMS 2000 subject classifications: 62F10; 62F15; 62F30

Keywords: Maximum likelihood estimator; Restricted parameter space; Squared error loss; Dominance; Simultaneous dominance; Robustness; Spherically symmetric distribution; Scale mixture of normals; Multivariate student distribution

*Corresponding author.
E-mail address: perronf@dms.umontreal.ca (F. Perron).

[1] Research supported by NSERC of Canada.

doi:10.1016/j.jmva.2003.09.009

1. Introduction

Consider the problem of estimating under squared error loss the location parameter θ of a spherically symmetric distribution, based on the observation X and with the constrained parameter space $\Theta(m) = \{\theta \in \mathbb{R}^p \colon ||\theta|| \leqslant m\}$ for some m fixed, $m > 0$. In the normal case with identity covariance matrix, Marchand and Perron [4] showed that, among classes of dominating estimators, the Bayes estimator δ_{BU} with respect to the boundary uniform prior on $\partial\Theta(m)$ dominates the maximum likelihood estimator δ_{mle} whenever $m \leqslant \sqrt{p}$. An interesting question is to investigate whether similar dominance results hold for other spherically symmetric distributions. This is indeed the objective of our research, and our results are focussed on: (i) the multivariate student distribution which represents perhaps one of the most important alternatives to the normal model, and (ii) on scale mixtures of normals for which the scaling random variable is bounded below by some positive constant with probability one.

The starting point of our inquiry is a sufficient condition (Theorem 1) for an estimator to dominate δ_{mle}, which was implicitly given by Marchand and Perron [4, Theorem 3], and which is applicable in general to spherically symmetric distributions. In Section 3.2, we provide explicit dominance conditions applicable to multivariate student distributions. We also study how these conditions apply to δ_{BU}, and establish (Example 1) that the condition $m \leqslant \sqrt{p}$ is, whenever $d \geqslant p$, once again sufficient for δ_{BU} to dominate δ_{mle}. The common sufficient condition is interesting and somewhat surprising, in view of its simplicity, and the fact that both the functional form of the estimator δ_{BU} and the distribution under which the risks are evaluated vary with d.

We also can view the sufficient condition for dominance of Theorem 1 as a sufficient condition for simultaneous dominance (Theorem 2), meaning a condition under which a single estimator δ_0 dominates δ_{mle} simultaneously for a subclass of spherical distributions. Of course, it is the hope that such a simultaneous condition of dominance can be made explicit for important subclasses of spherical distributions, possibly including the normal case. Simultaneous dominance is an appealing property in view of the intrinsic motivation of assessing or searching for procedures that retain good or optimal properties over a range of probability models. Although there seems to be a relative paucity of results in this direction, this is not a new theme; for instance, a fair amount of work on estimating a multivariate mean (without constraints) has dealt with procedures that are robust, in the sense that they perform well not only for the normal model, but also for a range of spherical or elliptical models. As a first example, Cellier and Fourdrinier [1], gave a class of estimators that dominate the unbiased estimator, for $p \geqslant 3$, simultaneously for all spherically symmetric distributions subject to (weak) risk finiteness conditions. A second example is given by the work of Srivastava and Bilodeau [7] who demonstrate robust dominance properties of the Stein estimator (in dominating the unbiased estimator) for elliptical distributions which are scale mixtures of normals.

In Section 3.3, we focus again on multivariate student distributions with d degrees of freedom and obtain two examples of simultaneous dominance. In particular, we obtain an explicit estimator δ_0 which, for $m < \sqrt{p}$, dominates δ_{mle} simultaneously for all multivariate student distributions with $d \geqslant p$ as well as the normal distribution with identity covariance matrix. This is a particular interesting result since no theoretical elements that we know of guaranteed the existence of such a simultaneously dominating δ_0. The simultaneous dominating estimators obtained, although simple, may well fail to be attractive for a given single distribution, but the result permits us to envisage locally (or globally) more attractive estimators to enjoy the same simultaneous dominating property.

Finally, in Section 3.4, we provide conditions of simultaneous dominance for the subclass of scale mixtures of normals for which the scaling random variable is bounded below by some positive constant with probability one. Namely, Corollary 1 gives a robustness dominance property of a normal model boundary uniform Bayes estimator.

Before proceeding in Section 3 with these dominance results, we pursue with by collecting some further notations, definitions and properties for later use.

2. Definitions and preliminaries

Throughout, we shall denote $||x||$ and $||\theta||$ by r and λ, respectively. We consider distributions with probability density functions

$$f_\theta(x) = h(||x - \theta||), \tag{1}$$

where h is such that $h(t) < h(0)$ for all $t > 0$. For such distributions, the maximum likelihood estimator of θ is uniquely given by $\delta_{\text{mle}}(x) = (\frac{m}{||x||} \wedge 1)x$. The function $g_{h,\lambda}(r) = E_\theta[\frac{\theta' X}{||X||} \mid ||X|| = r]$ plays a pivotal role in our dominance results as it intervenes in both (i) the decomposition of risks (see Theorem 1), and (ii) the functional form of the Bayes estimator $\delta_{g_{h,\lambda}}$ with respect to a uniform prior on the sphere $\{\theta\colon ||\theta|| = \lambda\}$, given by (e.g., [3, proof of Theorem 2.3])

$$\delta_{g_{h,\lambda}}(x) = \frac{1}{r} g_{h,\lambda}(r) x.$$

Of particular interest is the boundary uniform prior, and the associated Bayes estimator $\delta_{\text{BU}} = \delta_{g_{h,m}}$ which was shown by Marchand and Perron [4] to dominate δ_{mle} in the normal case with identity covariance matrix whenever $m \leqslant \sqrt{p}$.

We further define $\bar{g}_{h,m}(r) = \sup_{0 \leqslant \lambda \leqslant m} \{g_{h,\lambda}(r)\}$, and $A_{h,m} = \{r > 0\colon \bar{g}_{h,m}(r) < r\}$.

Remark 1. From its definition and the Cauchy–Schwarz inequality, it is easy to see that $g_{h,\lambda}(r) < \lambda \leqslant m$. Hence $A_{h,m}$ always contains the set $[m, \infty)$.

Our conditions for dominance in Section 3 below are given first implicitly in terms of $\bar{g}_{h,m}$ and $A_{h,m}$ (Theorems 1 and 2), and we proceed by developing more explicit conditions in the multivariate student case (Theorems 3 and 4). In order to achieve this, we require two technical lemmas. We begin with an expression for $g_{h,\lambda}$ for scale mixture of normals where X admits the representation

$$\mathscr{L}(X|V = v) = N_p(\theta, v^{-1}I_p) \tag{2}$$

for some positive random variable V.

Lemma 1 (Marchand [3]). *For scale mixture of normals as defined above, we have*

$$g_{h,\lambda}(r) = \lambda \frac{E[I_{\frac{p}{2}}(tV)e^{-sV}V]}{E[I_{\frac{p}{2}-1}(tV)e^{-sV}V]},$$

where $t = \lambda r$, $s = (\lambda^2 + r^2)/2$, *and* $I_v(y)$; $v \geqslant -\frac{1}{2}$, $y \geqslant 0$; *is the modified Bessel function of order* v *given by* $I_v(y) = \sum_{i\geqslant 0} \frac{(\frac{y}{2})^{v+2i}}{i!\Gamma(i+v+1)}$.

Remark 2. When referring to a specific distribution of the mixing parameter in (2) for which $E[V] < \infty$, we can assume without loss of generality (and we will hereafter) that $E[V] = 1$. This is so since, whenever $E[V] \neq 1$, we can always transform the problem to work with: (i) the observation $X^* = \sqrt{E[V]}X$, (ii) the representation for X^* as in (2) (with location parameter $\theta^* = \theta\sqrt{E[V]}$) corresponding to the mixing parameter $V^* =^d \frac{V}{E[V]}$, for which $E[V^*] = 1$, and (iii) the constraint $||\theta^*|| \leqslant m^*$ with $m^* = m\sqrt{E[V]}$.

We now continue with some useful properties of $g_{h,\lambda}(r)$ for the cases in (2) where $\mathscr{L}(V) = \mathrm{Gamma}(a, b)$. As mentioned in Remark 2, there is no loss of generality in limiting ourselves to the cases where $E[V] = 1$; i.e., $a = b$; which corresponds to the multivariate student cases with degrees of freedom $d, d > 0$, with $a = b = \frac{d}{2}$.

Lemma 2. *If the distribution of* X *follows a multivariate student distribution with* d *degrees of freedom and* $m \leqslant \sqrt{d}$ *then*

$$\bar{g}_{h,m}(r) = g_{h,m}(r) \tag{3}$$

$$\leqslant \frac{m^2 r}{m^2 + r^2 + d}\left[1 + \left(1 \vee \frac{d}{p}\right)\right] \tag{4}$$

for all $r > 0$.

Proof. See the appendix.

3. Dominance results

We begin this section with a sufficient condition for an estimator $\delta_g(x) = \frac{1}{r} g(r)x$ to dominate δ_{mle}. The proof is essentially the same as the one given by Marchand and Perron [4] in the normal case, but given here for sake of completeness.

3.1. General dominance results

Theorem 1. *For distributions as in* (1) *and* $\delta_g(x) = \frac{1}{r} g(r)x$, *the estimator* δ_g *dominates* δ_{mle} *as long as*

$$2\bar{g}_{h,m}(r) - (r \wedge m) < g(r) < (r \wedge m)$$

for all $r \in A_{h,m}$ *and* $g(r) = r$ *otherwise.*

Proof. We have

$$\begin{aligned} R(\theta, \delta_g) &= E_\theta\left[||g(||X||)\frac{X}{||X||} - \theta||^2 \right] \\ &= E_\theta\left[||\theta||^2 + g^2(||X||) - 2g(||X||)\frac{\theta' X}{||X||} \right] \\ &= ||\theta||^2 + E_\theta[\{g(||X||) - g_{h,\lambda}(||X||)\}^2 - g^2_{h,\lambda}(||X||)]. \end{aligned}$$

Hence, $R(\theta, \delta_{\text{mle}}) - R(\theta, \delta_g) = E_\theta[\{g_{\text{mle}}(||X||) - g(||X||)\}\{g_{\text{mle}}(||X||) + g(||X||) - 2g_{h,\lambda}(||X||)\}]$, which is indeed positive for all $\theta \in \Theta(m)$ under the stated conditions. □

Remark 3. Note that $\delta_{\bar{g}_{h,m}}$ satisfies the conditions of Theorem 1 whenever $A_{h,m} = (0, \infty)$, while its truncated version (i.e., with $g(r) = \bar{g}_{h,m}(r) \wedge g_{\text{mle}}(r)$) always dominates δ_{mle}. In the normal case with identity covariance matrix (i.e., [4]), it was established that (i) $\delta_{\bar{g}_{h,m}} = \delta_{\text{BU}}$; and that (ii) $A_{h,m} = (0, \infty)$ if and only if $m \leqslant \sqrt{p}$. Now, by requiring an estimator δ_g to fulfill the conditions of Theorem 1 for all h in a family $\mathscr{H}$ of distributions, we obtain the following simultaneous dominance result. Applications of Theorem 2 to (i) multivariate student distributions and to (ii) scale mixtures of normals for which the scaling random variable is bounded from below are presented in Sections 3.3 and 3.4.

Theorem 2. *Let* $\delta_g(x) = \frac{1}{r} g(r)x$. *The estimator* δ_g *dominates* δ_{mle} *simultaneously for all* $h \in \mathscr{H}$ *as long as*

$$2 \sup_{h \in \mathscr{H}} \bar{g}_{h,m}(r) - (r \wedge m) < g(r) < (r \wedge m)$$

on the set $A_{\mathscr{H},m}$, *and* $g(r) = r$ *otherwise, with* $A_{\mathscr{H},m} = \{r\colon \ \sup_{h \in \mathscr{H}} \bar{g}_{h,m}(r) < (r \wedge m)\}$.

3.2. Dominance results for the multivariate student distribution

In this section, Theorem 2 is an application of the dominance results of Theorem 1 to the case of a multivariate student distribution. Moreover, its part (d) applies in particular to δ_{BU} (see Example 1) and follows from the dominance conditions: $A_{h,m} = (0, \infty)$ and $\bar{g}_{h,m} = g_{h,m}$, mentioned in Remark 3.

Theorem 3. *Assume that the distribution of X follows a multivariate student distribution with d degrees of freedom and $\delta_g(x) = \frac{1}{r} g(r)x$.*

(a) *If $m \leqslant \sqrt{p \wedge d}$, $d < \{1 + 2[(p/m^2 - 1) + \sqrt{(p/m^2 - 1)^2 + p/m^2}]\}p$, and g is such that*

$$2\frac{m^2 r}{m^2 + r^2 + d}\left[1 + \left(1 \vee \frac{d}{p}\right)\right] - (r \wedge m) < g(r) < (r \wedge m)$$

for all $r > 0$, then δ_g dominates δ_{mle}.

(b) *If $m \leqslant \sqrt{p}$, $d \geqslant \{1 + 2[(p/m^2 - 1) + \sqrt{(p/m^2 - 1)^2 + p/m^2}]\}p$, and g is such that $g(r) = m$ if $\{2r - m(1 + d/p)\}^2 \geqslant m^2(1 + d/p)^2 - 4(m^2 + d)$ and*

$$2\frac{m^2 r}{m^2 + r^2 + d}\left(1 + \frac{d}{p}\right) - (r \wedge m) < g(r) < (r \wedge m)$$

otherwise, then δ_g dominates δ_{mle}.

(c) *If $m \leqslant \sqrt{d}$ and g is such that*

$$2g_{h,m}(r) - (r \wedge m) < g(r) < (r \wedge m)$$

for all $r \in A_{h,m}$ and $g(r) = r$ otherwise, then δ_g dominates δ_{mle}.

(d) *If $m \leqslant \sqrt{p \wedge d}$ and g is such that*

$$2g_{h,m}(r) - (r \wedge m) < g(r) < (r \wedge m)$$

for all $r > 0$, then δ_g dominates δ_{mle}.

Proof. (a) Here is an application of Theorem 1 and Lemma 2. We need only to verify that $\frac{m^2 r}{m^2+r^2+d}[1 + (1 \vee \frac{d}{p})] < (r \wedge m)$ for all $r > 0$. This is done by verifying that: (i) $\frac{m^2 r}{m^2+r^2+d}[1 + (1 \vee \frac{d}{p})] < r$ for all $r > 0$ if and only if $m \leqslant \sqrt{p \wedge d}$; and (ii) $\frac{m^2 r}{m^2+r^2+d}[1 + (1 \vee \frac{d}{p})] < m$ for all $r > 0$ if and only if $d < \{1 + 2[(p/m^2 - 1) + 2\sqrt{(p/m^2 - 1)^2 + p/m^2}]\}p$.

(b) This proof is similar to the one of part (a) except that here $\frac{m^2 r}{m^2+r^2+d}(1 + \frac{d}{p}) \leqslant m$ whenever $\{2r - m(1 + d/p)\}^2 \geqslant m^2(1 + d/p)^2 - 4(m^2 + d)$.

(c) This is a direct application of Theorem 1 and Lemma 2.

(d) Given the result in part (c), we need only to verify that $A_{h,m} = (0, \infty)$. From Lemma 2, we know that $g_{h,m}(r) \leqslant \frac{m^2 r}{m^2+r^2+d}[1 + (1 \vee \frac{d}{p})]$, while the proof of part (a) tells us that $\frac{m^2 r}{m^2+r^2+d}[1 + (1 \vee \frac{d}{p})] < r$ for all $r > 0$ if $m \leqslant \sqrt{p \wedge d}$, which yields the result. □

Example 1. The estimator δ_{BU} dominates δ_{mle} whenever $m \leqslant \sqrt{p \wedge d}$. In fact, this is a special case of Theorem 3, part (d).

Remark 4. For $d \geqslant p$, δ_{BU} dominates δ_{mle} whenever $m \leqslant \sqrt{p}$, duplicating Marchand and Perron's [4] sufficient condition in the normal case. Finally, it also can be shown that for $d \geqslant p$, the condition $m \leqslant \sqrt{p}$ is also necessary for $A_{h,m}$ to equal $(0, \infty)$. This is established by considering the necessary condition $\lim_{r \to 0} \frac{\bar{g}_{h,m}(r)}{r} \leqslant 1$, and using expression (8) in the proof of Lemma 2 (see the appendix) to infer that $\lim_{r \to 0} \frac{\bar{g}_{h,m}(r)}{r} = \frac{m^2}{m^2+d} \frac{d+p}{p}$.

3.3. Simultaneous dominance results for multivariate student distributions

We now turn to applications of Theorem 2, that is the specification of estimators that dominate δ_{mle} for several distributions simultaneously, and results are given herein for multivariate student distributions. Note that the choice $g(r) = \sup_{h \in \mathscr{H}} \bar{g}_{h,m}(r)$ for $r \in A_{\mathscr{H},m}$ satisfies the conditions of Theorem 2, while the above results imply that $A_{\mathscr{H},m} = (0, \infty)$ for $m \leqslant \sqrt{p}$ with $\mathscr{H}$ being the multivariate student family with degrees of freedom $d \geqslant p$. However, this estimator is not given explicitly. The results below pertaining to the family of multivariate student distributions with d degrees of freedom, $d \geqslant p$, are of particular interest since (i) dominance is shown to hold as well for the normal distribution with identity covariance matrix, and (ii) the family includes all univariate student distributions with d degrees of freedoms, $d \geqslant 1$, whenever $p = 1$.

Theorem 4. *Assume that the distribution of X follows a multivariate student distribution with d degrees of freedom and $\delta_g(x) = \frac{1}{r} g(r)x$.*

(a) *If $m \leqslant \sqrt{d_0}$, $d_0 \leqslant p$, and g it is such that*

$$4 \frac{m^2 r}{m^2 + r^2 + d_0} - (r \wedge m) < g(r) < (r \wedge m)$$

for all $r > 0$ then δ_g dominates δ_{mle} for all d, $d_0 \leqslant d \leqslant p$.

(b) *If $m < \sqrt{p}$, $d_0 \geqslant p$, and g is such that*

$$2 \frac{m^2 r}{p} \left(1 \vee \frac{p + d_0}{m^2 + r^2 + d_0} \right) - (r \wedge m) < g(r) < (r \wedge m)$$

for all $0 < r < p/m$, and $g(r) = m$ otherwise, then δ_g dominates δ_{mle} for all d, $d \geqslant d_0$, and for the normal distribution case as well.

Proof. The proof is an application of Theorems 2 and 3.

(a) Since $mr^2/(m^2 + r^2 + d)$ is a decreasing expression in d, our result satisfies the conditions of Theorem 3, part (a), for all d, $d_0 \leqslant d \leqslant p$.

(b) We use the results of Theorem 3, part (b). Since $\frac{m^2 r}{p}(1 \vee \frac{p+d_0}{m^2+r^2+d_0}) < (r \wedge m)$ for all $0 < r < p/m$, and $\frac{m^2 r(1+\frac{d}{p})}{m^2+r^2+d} \leqslant \frac{m^2 r}{p}(1 \vee \frac{p+d_0}{m^2+r^2+d_0})$ for all $r > 0$, $d \geqslant d_0$ we obtain our result. Finally, δ_g and δ_{mle} are both bounded and the densities converge to the one of a normal distribution as $d \to \infty$ which implies that the risk functions converge and our result is still valid for the normal distribution. $\square$

Example 2. Translating directly the conditions of Theorem 4b to the univariate case with $d_0 = 1$, we obtain for $m < 1$ that the estimator

$$\delta_0(x) = m\left\{m\left(1 \vee \frac{2}{m^2 + x^2 + 1}\right) \wedge \frac{1}{|x|}\right\}x$$

dominates δ_{mle} simultaneously for all student distributions with degrees of freedom $d \geqslant 1$, and the normal distribution as well.

3.4. Simultaneous dominance results for scale mixtures of normals for which the mixing random variable V is bounded above

As in Section 3.3, we give below applications of Theorem 2 to the subclass of scale mixtures of normals, as in (2), for which the mixing random variable V is bounded above by $V_{\max}$ (say); $V_{\max} < \infty$; with probability 1. We will denote this subclass $\mathscr{H}_{V_{\max}}$. Also, observe that the subclass $\mathscr{H}_{V_{\max}}$ includes normal $N_p(\theta, \sigma^2 I_p)$ distributions for which $\sigma^2 \geqslant (V_{\max})^{-1}$. Analogously to the development above for multivariate student distributions, we make use of an upper bound for $\sup_{h \in \mathscr{H}_{V_{\max}}} \bar{g}_{h,m}(r)$.

Lemma 3. *We have*

$$\sup_{h \in \mathscr{H}_{V_{\max}}} \bar{g}_{h,m}(r) \leqslant m\rho_{p/2-1}(mrV_{\max}),$$

where, for $v \geqslant -\frac{1}{2}, t > 0,\ \rho_v(t) = \frac{I_{v+1}(t)}{I_v(t)}$.

Proof. From Lemma 1, we may write for a scale mixture of normals $g_{h,\lambda}(r) = \lambda E[\rho_{p/2-1}(\lambda r W)]$ with W a random variable with density proportional to $I_{p/2-1}(tw)e^{-sw}w\,dG(w)$; G being the cdf of V. Now, since $\rho_{p/2-1}(tw)$ is an increasing and concave function of w; $w > 0$; (e.g., [6,8]) we have by Jensen's inequality for $h \in \mathscr{H}_{V_{\max}}$

$$g_{h,\lambda}(r) \leqslant \lambda\rho_{p/2-1}(\lambda r E(W)) \leqslant \lambda\rho_{p/2-1}(\lambda r V_{\max}),$$

given that W is bounded above by $V_{\max}$ with probability 1. Finally, by using again the increasing property of ρ and the fact that $\rho \geqslant 0$, we obtain $\bar{g}_{h,m}(r) \leqslant m\rho_{p/2-1}(mrV_{\max})$ for all $h \in \mathscr{H}_{V_{\max}}$, which yields the result. $\square$

We now are ready to pursue with the following simultaneous dominance results.

Theorem 5. *For the subclass of scale mixtures $\mathscr{H}_{V_{\max}}$, an estimator $\delta_g(x) = \frac{1}{r}g(r)x$ dominates δ_{mle} simultaneously whenever*

$$2\rho_{p/2-1}(mrV_{\max}) - (r \wedge m) < g(r) < (r \wedge m),$$

on the set $\{r\colon m\rho_{p/2-1}(mrV_{\max}) < r\}$, and $g(r) = r$ otherwise.

Proof. This is a direct consequence of Theorem 2 and Lemma 3.

Observe that the truncated version of a normal model boundary uniform Bayes estimator δ_{BU} (i.e., the estimator δ_g with $g(r) = (m\rho_{p/2-1}(mrV_{\max}) \wedge r)$ always satisfies the conditions of the Theorem (note that $\rho < 1$), while as now stated, if truncation is not necessary the simultaneous dominance property applies to δ_{BU}.

Corollary 1. *The estimator δ_{BU} associated with the model $N_p(\theta, (V_{\max})^{-1}I_p)$ dominates δ_{mle} on $\Theta(m)$ simultaneously for all $h \in \mathscr{H}_{V_{\max}}$ whenever $m \leqslant (V_{\max})^{-1}\sqrt{p}$.*

Proof. The result follows from Theorem 5, and the fact that $m\rho_{p/2-1}(mrV_{\max}) < r$; $r > 0$; whenever $mV_{\max} \leqslant \sqrt{p}$ (see [4]; or Remark 3). $\square$

This last example is particularly interesting since the dominating δ_{BU} possesses nice properties for the model from which it is derived (i.e., Bayes, admissible, minimax for small enough values of m which include values of $m \leqslant (V_{\max})^{-1}\sqrt{p}$ as shown by Marchand and Perron, [5]), and at the same time is robust in its dominance of δ_{mle} to departures of the underlying distribution to members of $\mathscr{H}_{V_{\max}}$. Ongoing work of ours is aimed at extending this robustness result to other scale mixtures of normals, and to other spherical distributions.

Acknowledgments

The authors thank two anonymous reviewers for their helpful comments and suggestions.

Appendix A

In the appendix we shall prove that expressions (3) and (4) of Lemma 2 are valid whenever X follows a multivariate student distribution with d degrees of freedom and $m \leqslant \sqrt{d}$.

Proof of Expression (3) in Lemma 2. Let $T = \theta' X/\lambda R$ so $g_{h,\lambda}(r) = \lambda E_\lambda[T|R^2 = r^2]$. If we can show that $E_\lambda[T|R^2 = r^2]$ is nondecreasing in λ for all r, $r > 0$ then we shall have $\bar{g}_{h,m} = g_{h,m}$, given that $g_{h,\lambda} \geqslant 0$ as seen by Lemma 1. Moreover, the expression $E_\lambda[T|R^2 = r^2]$ will be nondecreasing in λ for all r, $r > 0$ if the conditional distribution of T given that $R^2 = r^2$ has monotone increasing likelihood ratio in T for all $r > 0$, where λ is the parameter and r is fixed. Let $f_\theta(x) = h(||x-\theta||)$ as in expression (1).

If $p = 1$, then T is a discrete random variable taking the values -1 and 1, R^2 is a continuous random variable on $(0, \infty)$, and their likelihood is given by $\varphi_{\lambda,1}$ with

$$\varphi_{\lambda,1}(t, r^2) = \frac{1}{2r} h(\sqrt{r^2 + \lambda^2 - 2\lambda rt}).$$

Similarly, if $p > 1$ then T, R^2 have a joint density $\varphi_{\lambda,p}$ on $(-1,1) \times (0, \infty)$, which has been obtained by Kariya and Eaton [2], and it is given by

$$\varphi_{\lambda,p}(t, r^2) = 2\frac{(\sqrt{\pi})^{p-1}}{\Gamma((p-1)/2)} r^{p-2}(1-t^2)^{(p-3)/2} h(\sqrt{r^2 + \lambda^2 - 2\lambda rt}).$$

In any case, the monotone likelihood property will hold if we can show that the derivative, with respect to t, of the expression $\log(h(\sqrt{r^2+\lambda^2-2\lambda rt}))$ is nondecreasing in λ for all $t \in [-1,1]$, $\lambda > 0$. We now return to the multivariate student distribution set up, that is h is given by $h(z) = (2\pi)^{-p/2} E[V^{p/2}\exp(-z^2V/2)]$ with $\mathscr{L}(V) = \text{Gamma}(d/2, d/2)$. We obtain

$$\begin{aligned}\frac{\partial}{\partial t}\log(h(\sqrt{r^2+\lambda^2-2\lambda rt})) &= \frac{\frac{\partial}{\partial t} E[V^{p/2}\exp(-\{r^2+\lambda^2-2\lambda rt\}V/2)]}{E[V^{p/2}\exp(-\{r^2+\lambda^2-2\lambda rt\}V/2)]}\\ &= \frac{\frac{\partial}{\partial t}\{r^2+\lambda^2-2\lambda rt+d\}^{-(d+p)/2}}{\{r^2+\lambda^2-2\lambda rt+d\}^{-(d+p)/2}}\\ &= \frac{(p+d)\lambda r}{(r^2+\lambda^2-2\lambda rt+d)}\end{aligned}$$

and the last expression is increasing in λ on $[0, \sqrt{d}]$ for all $r > 0$, $t \in [-1,1]$. □

Proof of expression (4) in Lemma 2. From Lemma 1, it follows that

$$\begin{aligned}g_{h,\lambda}(r) &= \lambda\frac{\int_0^\infty I_{\frac{p}{2}}(tv)v^{\frac{d}{2}}e^{-(\frac{d}{2}+s)v}\,dv}{\int_0^\infty I_{\frac{p}{2}-1}(tv)v^{\frac{d}{2}}e^{-(\frac{d}{2}+s)v}\,dv}\\ &= \lambda\frac{\int_0^\infty I_{\frac{p}{2}}(x)x^{\frac{d}{2}}e^{-\frac{x}{u}}\,dx}{\int_0^\infty I_{\frac{p}{2}-1}(x)x^{\frac{d}{2}}e^{-\frac{x}{u}}\,dx},\end{aligned} \tag{A.1}$$

with the change of variables $x = tv$ and $u = 2\lambda r/(\lambda^2 + r^2 + d)$. Now, by expanding $I_v(x)$ and interchanging sum and integral, we obtain

$$\begin{aligned}
\int_0^\infty I_v(x) x^{\frac{d}{2}} e^{-\frac{x}{u}}\,dx &= \sum_{i\geqslant 0} \frac{(\frac{1}{2})^{2i+v}}{i!\Gamma(i+v+1)} \int_0^\infty x^{v+\frac{d}{2}+2i} e^{-\frac{x}{u}}\,dx \\
&= \left(\frac{1}{2}\right)^v \sum_{i\geqslant 0} \frac{(\frac{1}{2})^{2i}}{i!\Gamma(v+1)} \frac{\Gamma(v+\frac{d}{2}+1+2i) u^{v+\frac{d}{2}+2i+1}}{(v+1)_i} \\
&= \frac{(\frac{1}{2})^v u^{v+\frac{d}{2}+1}}{\Gamma(v+1)} \sum_{i\geqslant 0} \frac{u^{2i}}{i!} \frac{\Gamma(v+\frac{d}{2}+1)(\frac{v+\frac{d}{2}+1}{2})_i(\frac{v+\frac{d}{2}+2}{2})_i}{(v+1)_i} \\
&= \frac{\Gamma(v+\frac{d}{2}+1)}{\Gamma(v+1)} \left(\frac{1}{2}\right)^v u^{v+\frac{d}{2}+1} \\
&\quad \times {}_2F_1\left(\frac{v+\frac{d}{2}+1}{2}, \frac{v+\frac{d}{2}+2}{2}; v+1; u^2\right) \qquad \text{(A.2)}
\end{aligned}$$

with ${}_2F_1(a_1, a_2; a_3; z) = \sum_{i\geqslant 0} \frac{(a_1)_i (a_2)_i}{(a_3)_i} \frac{z^i}{i!}$; and $(d)_i = \frac{\Gamma(d+i)}{\Gamma(d)}$. By using standard operations on hypergeometric functions, for any a_1, a_2, a_3 with $a_3 > 0$, we obtain that

$$\begin{aligned}
{}_2F_1(a_1, a_2+1; a_3+1; z) &= \sum_{i\geqslant 0} \frac{(a_1)_i (a_2+1)_i}{(a_3+1)_i} \frac{z^i}{i!} \\
&= \sum_{i\geqslant 0} \frac{(a_1)_i (a_2)_i \frac{(a_2+i)}{a_2}}{(a_3)_i \frac{(a_3+i)}{a_3}} \frac{z^i}{i!} \\
&= \frac{a_3}{a_2} \sum_{i\geqslant 0} \frac{(a_1)_i (a_2)_i}{(a_3)_i} \frac{(a_3+i)+(a_2-a_3)}{a_3+i} \frac{z^i}{i!} \\
&= \frac{a_3}{a_2} \Big[{}_2F_1(a_1, a_2; a_3; z) \\
&\quad + \frac{a_2 - a_3}{a_3} {}_2F_1(a_1, a_2; a_3+1; z) \Big]. \qquad \text{(A.3)}
\end{aligned}$$

Combining expressions (A.1)–(A.3), with $v = p/2 - 1$ and $v = p/2$, $a_1 = (p + d + 2)/4$, $a_2 = (p + d)/4$ and $a_3 = p/2$, we may write

$$g_{h,\lambda}(r) = \frac{\lambda^2 r}{\lambda^2 + r^2 + d} \left\{ 2 + \frac{(d-p)}{p} E_u \left[\frac{p/2}{p/2 + Y} \right] \right\}, \qquad \text{(A.4)}$$

where Y is a discrete random variable having probability mass function p_u with

$$p_z(y) \propto \frac{(a_1)_y (a_2)_y}{(a_3)_y} \frac{z^{2y}}{y!}, \quad y = 0, 1, \ldots$$

for $0<z<1$. Since $2+\frac{(d-p)}{p}E_u[\frac{p/2}{p/2+Y}]\leqslant 1+(1\vee\frac{d}{p})$, we obtain that

$$g_{h,\lambda}(r)\leqslant\frac{\lambda^2 r}{\lambda^2+r^2+d}\left\{1+\left(1\vee\frac{d}{p}\right)\right\}$$

for all $\lambda,\ 0\leqslant\lambda\leqslant m$. Setting $\lambda=m$ leads to the conclusion. $\square$

References

[1] D. Cellier, D. Fourdrinier, Shrinkage estimators under spherical symmetry for the general linear model, J. Multivariate Anal. 52 (1995) 338–351.

[2] T. Kariya, M. Eaton, Robust tests for spherical symmetry, Ann. Statist. 5 (1977) 206–215.

[3] É. Marchand, Estimation of a multivariate mean with constraints on the norm, Canad. J. Statist. 21 (1993) 359–366.

[4] É. Marchand, F. Perron, Improving on the MLE of a bounded normal mean, Ann. Statist. 29 (2001) 1078–1093.

[5] É. Marchand, F. Perron, On the minimax estimator of a bounded normal mean, Statist. Probab. Lett. 58 (2002) 327–333.

[6] C. Robert, Modified Bessel functions and their applications in probability and statistics, Statist. Probab. Lett. 9 (1990) 155–161.

[7] M.S. Srivastava, M. Bilodeau, Stein estimation under elliptical distributions, J. Multivariate Anal. 28 (1989) 247–259.

[8] G.S. Watson, Statistics on Spheres, Wiley, New York, 1983.